

Cyber bullying Detection on Social Networks Using Machine Learning Approaches

MOHAMMED LIJAS C

MES20MCA-2028

PRODUCT OWNER: MR VASUDEVAN T.V

TABLE OF CONTENTS

- 1. Introduction**
- 2. Modules**
- 3. Data Flow Diagram**
- 4. Developing Environment**
- 5. Product Backlog**
- 6. User Stories**
- 7. Project Plan**
- 8. Sprint Plans**
- 9. Sprint Actual**

INTRODUCTION

- Cyberbullying can be defined as aggressive, intentional actions performed by an individual or a group of people via digital communication methods such as sending messages and posting comments against a victim.
- Different from traditional bullying that usually occurs at school during face-to-face communication, cyber bullying on social media can take place anywhere at any time.
- Cyberbullying frequently leads to serious mental and physical distress, particularly for women and children, and even sometimes force them to attempt suicide.
- The purpose of this project is to design and develop an effective technique to detect online abusive and bullying messages by merging natural language processing and machine learning. Two distinct features, namely Bag-of - Words and term frequency-inverse text frequency (TF-IDF)
- In propped system, we suggest a cyberbullying detection model based on machine learning that can detect whether a text relates to cyberbullying or not.
- The results indicate that TF-IDF feature provides better accuracy than BoW. Users behaviour could be defined in several approaches like association rules in perspective of mining, complex graph activities, sequence mining etc. Suppose for two different user we have same behaviour we can recommend them each other.

METHODOLOGY

MACHINE LEARNING

Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning.

Term Frequency Inverse Document Frequency (TF IDF)

TF-IDF is frequently used in machine learning algorithms in various capacities, including stop-word removal. These are common words like “a, the, an, it” that occur frequently but hold little informational value. TF-IDF consists of two components, term frequency, and inverse document frequency

Term frequency can be determined by counting the number of occurrences of a term in a document.

IDF is calculated by dividing the total number of documents by the number of documents in the collection containing the term. It's useful for reducing the weight of terms that are common within a collection of documents.

FUTURE ENHANCEMENT

Current system that i implemented have only text filtering and block those kind of users.In Future i can implement image filtering which means those who post vulgar Post in form photos,comments etc.. and those users can be blocked by admin

MODULES

ADMIN

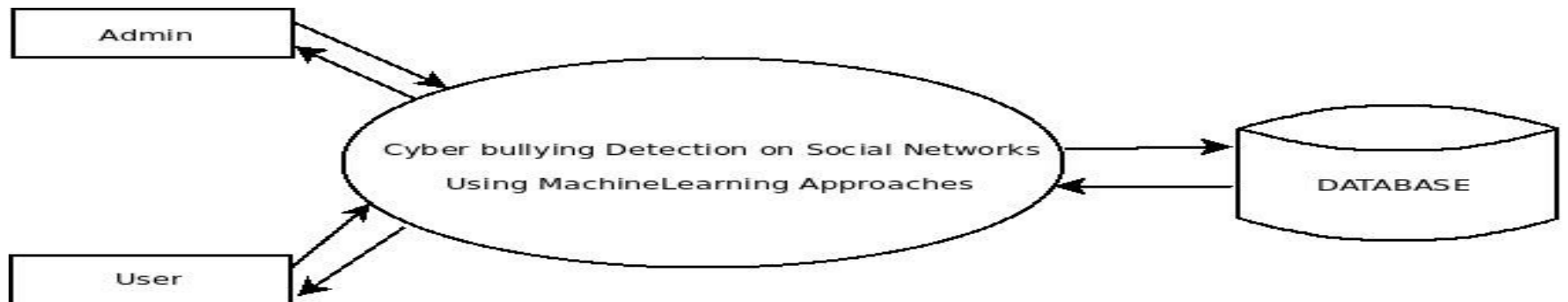
- 1.View user
- 2.Add bullying words
- 3.Add good words
- 4.View bullying words
- 5.View good words
- 6.View report

USER

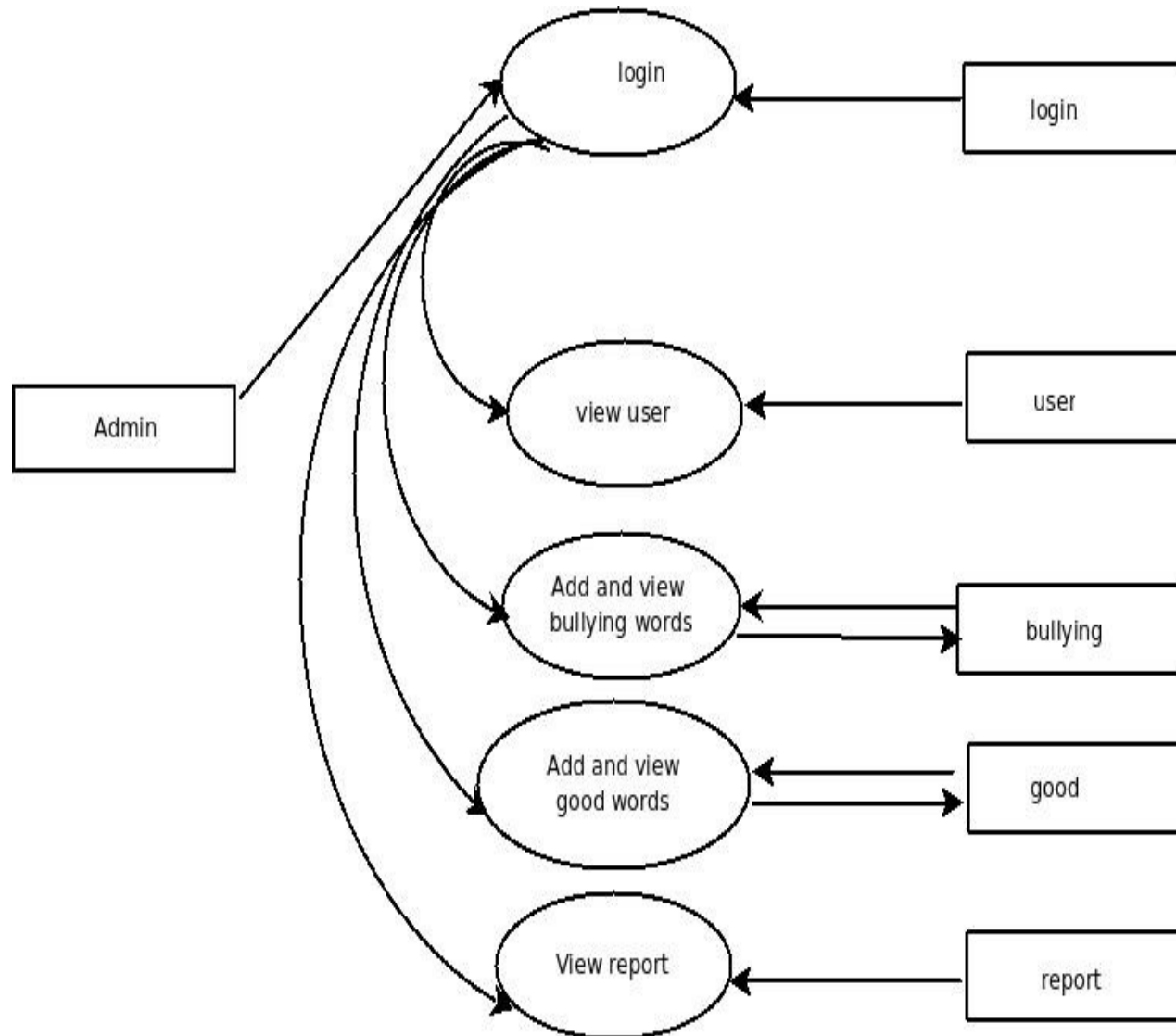
- 1.Registration
- 2.Add post
- 3.View my post
- 4.Chat
- 5.Add bullying words
- 6.Add friend request
- 7.View friend request
- 8.View recommendation

DATA FLOW DIAGRAM

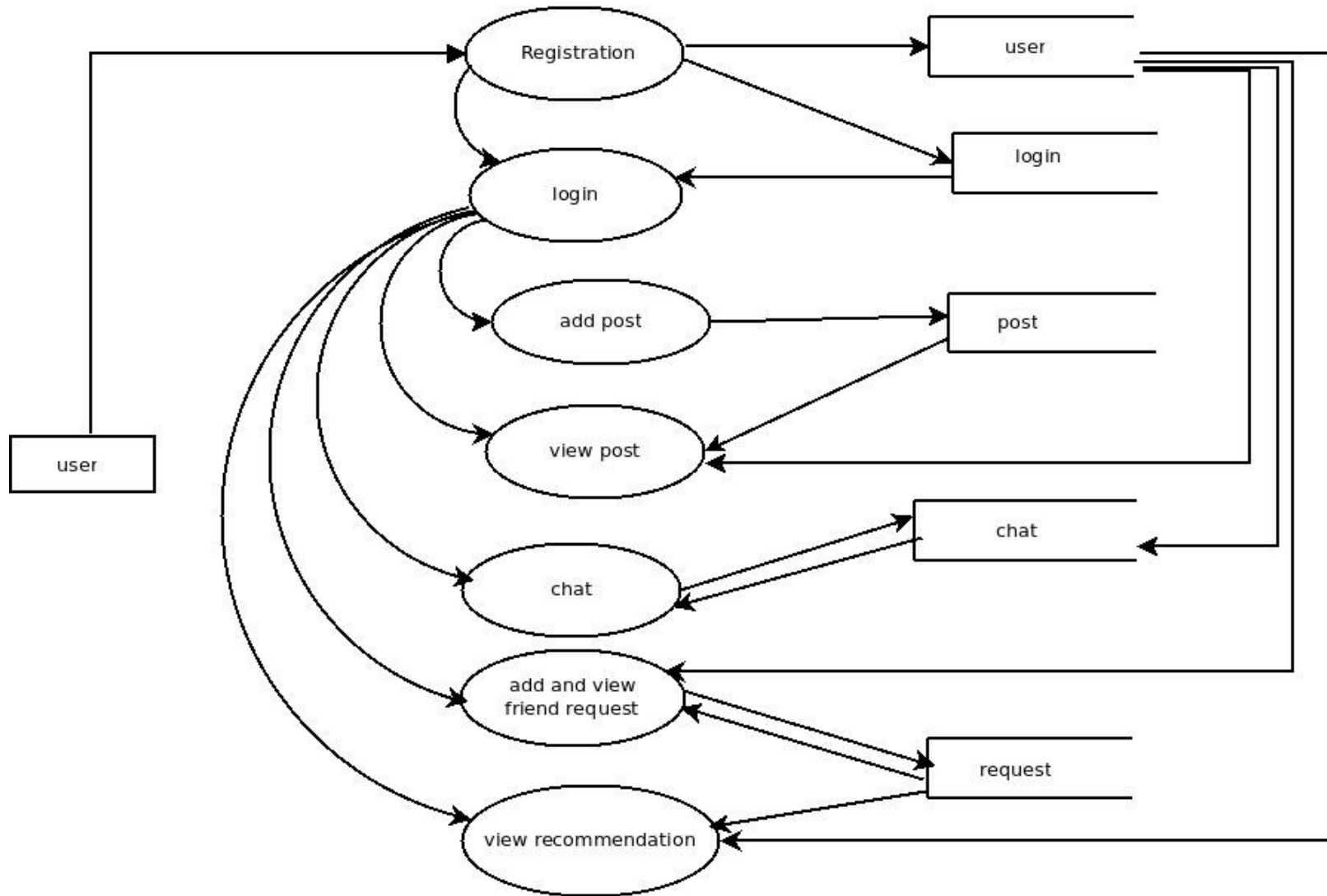
LEVEL 0



LEVEL 1.1



LEVEL 1.2



DEVELOPING ENVIRONMENT

Software Requirements

- Operating System - Windows 7 or Above , Linux
- Frontend – HTML,CSS,JavaScript
- Backend – Python ,MySQL
- Platform used - PyCharm, MySQL workbench
- Web Browser - Google Chrome, Fire fox, Microsoft Edge
- Frame work - Flask

PRODUCT BACKLOG

User Story ID	Priority <High/Medium/Low>	Size (Hours)	Sprint <#>	Status <Planned/In progress/Completed>	Release Date	Release Goal
1	Medium	2	1	Completed	08/01/2022	Table design
2	High	3		Completed	08/01/2022	Form design
3	High	5		Completed	08/01/2022	Basic coding
4	High	5	2	Completed	16/01/2022	Creation of data set
5	Medium	5		Completed		Prediction
6	High	5	3	Completed	27/01/2022	Filtering
7	high	5		Completed		Machine learning
8	Medium	5	4	Completed	10/02/2022	Testing data
9	High	5		Completed	19/02/2022	Output generation

USER STORIES

UserStoryID	As a <type of user>	I want to	So that I can
1	Admin	login	login successful with correct username and password
2	Admin	View users	View all user details
3	Admin	Add and view bullying words	Add and manage bullying words
4	Admin	Add and view good words	Add and manage good words for dataset
5	Admin	View report	View reports about bullying posts
6	User	Registration	Registration using personal information
7	User	Add post	Add a new post to OSN
8	User	View post	View all post OSN
9	User	Add and manage friend request	Accept or reject friend request
10	User	Chat	Chat with friends
11	User	View recommendation	View recommendation

PROJECT PLAN

User Story ID	Task Name	Start Date	End Date	Days	Status
1	Sprint 1	26/12/2021	28/12/2021	2	Completed
2		29/12/2021	31/12/2021	3	Completed
3		03/01/2022	08/01/2022	5	Completed
4	Sprint 2	09/01/2022	16/01/2022	8	Completed
5		18/01/2022	22/01/2022	5	Completed
6	Sprint 3	23/01/2022	27/01/2022	5	Completed
7		30/01/2022	05/02/2022	7	Completed
8	Sprint 4	06/02/2022	10/01/2022	5	Completed
9		16/02/2022	19/02/2022	4	Completed

SPRINT BACKLOG PLAN

Backlog Item	Status & completion date	Original estimate in hours	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14
User story #1,#2,#3		hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs
Table design	28/12/2021	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Form design	31/12/2021	3	0	0	0	1	1	1	0	0	0	0	0	0	0	0
Coding	08/01/2021	5	0	0	0	0	0	0	0	0	0	1	1	1	1	1
User story #4,#5																
Creation dataset	16/01/2022	5	1	1	0	1	0	1	0	1	0	0	0	0	0	0
Preprocessing	22/01/2022	5	0	0	0	0	0	0	0	0	0	1	1	1	1	1
User story #6,#7																
Training	27/01/2022	5	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Prediction	05/02/2022	5	0	0	0	0	0	0	0	1	0	1	1	1	0	1
User story #8,#9																
Testing data	10/02/2022	5	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Output	18/02/2022	5	0	0	0	0	0	0	0	0	0	0	1	1	1	1

SPRINT ACTUAL

Backlog Item	Status & completion date	Original estimate in hours	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14
User story #1,#2,#3		hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs	hrs
Table design	28/12/2021	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Form design	31/12/2021	3	0	0	0	1	1	1	0	0	0	0	0	0	0	0
Coding	08/01/2021	5	0	0	0	0	0	0	0	0	0	1	1	1	1	1
User story #4,#5																
Creation dataset	16/01/2022	5	1	1	0	1	0	1	0	1	0	0	0	0	0	0
Preprocessing	22/01/2022	5	0	0	0	0	0	0	0	0	0	1	1	1	1	1
User story #6,#7																
Training	27/01/2022	5	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Prediction	05/02/2022	5	0	0	0	0	0	0	0	1	0	1	1	1	0	1
User story #8,#9																
Testing data	10/02/2022	5	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Output generation	19/02/2022	5	0	0	0	0	0	0	0	0	0	2	1	1	1	1

THANK YOU